

VU Research Portal

Hebben wij eigenschappen nodig? : 'signs' en 'samples' in het psychologisch selectie-onderzoek

van der Flier, H.

1992

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

van der Flier, H. (1992). *Hebben wij eigenschappen nodig? : 'signs' en 'samples' in het psychologisch selectie-onderzoek*. Vrije Universiteit.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Dr. H. van der Flier

QB

08282

Hebben wij eigenschappen nodig?

**'Signs' en 'samples' in het
psychologisch selectie-onderzoek**

Rede uitgesproken bij de aanvaarding van het ambt van hoogleraar in de arbeids- en organisatiepsychologie aan de faculteit der psychologie en pedagogische wetenschappen van de Vrije Universiteit te Amsterdam op 3 april 1992.

vrije Universiteit amsterdam



Dr. H. van der Flier



Hebben wij eigenschappen nodig?

**'Signs' en 'samples' in het
psychologisch selectie-onderzoek**

vrije Universiteit amsterdam



Inleiding

Ik denk dat het de meesten van u weinig moeite zal kosten om een aantal psychologische eigenschappen op te sommen die van belang zijn voor het succesvol functioneren in arbeidsorganisaties. En anders hoeft u in de krant van zaterdag maar de pagina's met personeelsadvertenties op te slaan om een beeld te krijgen van het soort eigenschappen waaraan in dit verband gewoonlijk wordt gedacht. Veel gevraagd blijken personen (m/v) die zich kenmerken als dynamisch, flexibel en creatief, die beschikken over een goed analytisch vermogen, die stressbestendig zijn, sociaal vaardig, collegiaal en uiteraard ook klantgericht. Populair zijn daarnaast de eigenschappen zelfstandigheid, besluitvaardigheid, doorzettingsvermogen, verantwoordelijkheidsbesef en zo kunnen we nog wel even doorgaan.

Nu zullen we ons, bijvoorbeeld als het gaat om het beoordelen van sollicitanten voor een bepaalde functie toch wel enigszins moeten beperken, zowel wat betreft het aantal eigenschappen dat in aanmerking wordt genomen als wat betreft de m.b.t. die eigenschappen te stellen eisen. We zitten dus in ieder geval met twee vragen, t.w.:

1. Welke eigenschappen zijn voor de betreffende functie het meest wettelijk, en
2. Hoe krijgen we een beeld van de mate waarin sollicitanten deze eigenschappen bezitten.

Daar komen dan nog wel een aantal vragen achteraan, zoals hoe je de informatie over de verschillende eigenschappen weegt om tot geschiktheidsuitspraken of selectie-beslissingen te komen, maar daar wil ik het in deze voordracht verder niet over hebben. Bij de beantwoording van vragen als de zojuist genoemde worden nogal eens selectiepsychologen ingeschakeld. De vertaling van functie-eisen in psychologische eigenschappen en de meting van deze eigenschappen bij sollicitanten vormen dan ook belangrijke onderdelen van de psychologische adviespraktijk.

Toch heeft het gebruik van eigenschappen en eigenschapstermen binnen de selectie-psychologie nogal wat discussie opgeleverd, een discussie die voor verschillende typen instrumenten een wat verschillend karakter heeft gehad en die op dit ogenblik actueel is i.v.m. zogenaamde situatietests en assessment centers.

'Signs' versus 'samples'

Als startpunt neem ik een artikel van Wernimont en Campbell dat bijna 25 jaar geleden werd gepubliceerd (Wernimont en Campbell, 1968). In dit artikel worden twee manieren om arbeidsprestaties te voorspellen tegenover elkaar gezet, die worden aangeduid met de 'sign'-benadering en de 'sample'-benadering. De term 'sign' staat voor teken, signaal of aanduiding. De benadering houdt in dat psychologische tests worden opgevat als indicatoren van min of meer stabiele persoonseigenschappen. Verondersteld wordt dat tussen deze eigenschappen en gedrag in de arbeidssituatie bepaalde wetmatige (kans-) relaties bestaan. Kent men die relaties dan wordt het mogelijk om op grond van informatie over de betreffende eigenschappen (o.g.v. test-scores) voorspellingen te doen over het latere functioneren van kandidaten. De relatie tussen de voorspeller (de test) en de te voorspellen arbeidsprestatie loopt dus via het eigenschapsbegrip. Roe (1983, 1990) noemt deze benadering daarom de eigenschapsbenadering.

Tegenover deze 'sign'- of eigenschapsbenadering plaatsen Wernimont en Campbell de 'sample' (steekproef)- benadering. Uitgangspunt daarbij is dat toekomstig werkgedrag en toekomstige arbeidsprestaties het best voorspeld kunnen worden door te kijken naar gedrag in het verleden en naar eerdere prestaties in vergelijkbare situaties. Van het eigenschapsbegrip wordt bij deze benadering dus geen gebruik gemaakt, het gaat om een relatie tussen gedrag en gedrag. In de testsituatie wordt getracht aan de onderzochte gedrag te ontlokken dat overeenkomt met het gedrag dat in de arbeidssituatie van hem of haar gevraagd wordt en dat daar als het ware een steekproef uit vormt. Roe (ibid) spreekt in dit verband over de domein-steekproef-benadering of de gedragsgeneralisatie-benadering.

De reden dat Wernimont en Campbell met het formuleren van de tegenstelling 'sign versus sample' de gedragsgeneralisatie-benadering onder de aandacht brachten, was dat de toepassing van het klassieke validiteitsmodel dat sterk leunde op het eigenschapsbegrip toch minder goede resultaten te zien had gegeven dan men had gehoopt. Dit blijkt o.m. uit een twee jaar daarvoor, in 1966 gepubliceerd boek van Ghiselli met als titel 'The validity of occupational aptitude tests',

waarin een overzicht werd gegeven van de gemiddelde validiteiten van verschillende typen tests voor verschillende typen beroepen. Hoewel zeker kon worden geconstateerd dat bepaalde combinaties van tests en beroepen hogere validiteiten opleveren dan andere bleken gemiddelde validiteiten, ook bij gunstige combinaties, nauwelijks boven de .30 uit te komen. Volgens Wernimont en Campbell hing dit samen met het feit dat er in de selectiepsychologie ten onrechte het idee was ontstaan dat predictor en criterium inhoudelijk van elkaar zouden moeten verschillen. Volgens hen werkte deze notie contra-productief en zou juist gezocht moeten worden naar predictoren die zoveel mogelijk overeenkomst vertonen met het criterium of de criteria die voorspeld moeten worden. Dus eigenlijk zou je als voorspeller van latere arbeidsprestaties het liefst een stukje werkgedrag willen hebben.

Hoe stelden Wernimont en Campbell zich de selectie- of predictie-procedure volgens de gedragsgeneralisatie-benadering nu precies voor? Allereerst moet er volgens hen een grondige analyse van de betreffende functie plaatsvinden op grond waarvan prestatiedimensies worden geformuleerd, gedefinieerd door voorbeelden van effectief en ineffectief functiegedrag. Vervolgens wordt voor elke sollicitant informatie verzameld over voorafgaande werkervaring en opleiding om te bepalen welke van de relevante gedragingen of verrichtingen in het verleden van hem of haar zijn gevraagd en ook daadwerkelijk gedemonstreerd. Daarna wordt dan beoordeeld in welke mate de eerdere situaties lijken op de werksituatie en wordt de kans geschat dat de gedragingen zich in de werksituatie opnieuw zullen voordoen.

Na deze systematische analyse van achtergrondgegevens zou natuurlijk gevonden kunnen worden dat vereist functiegedrag zich in het verleden niet heeft voorgedaan of zich niet heeft kunnen voordoen. De achtergrondgegevens dienen in dat geval aangevuld te worden door m.b.v. arbeidsproeven en functie-simulaties te verzamelen relevant gedrag. Tenslotte kunnen als aanvulling hierop ook nog meer traditionele capaciteitentests worden afgenomen, maar dan vooral gericht op specifieke cognitieve functies.

Vergelijken we deze procedure met die vanuit de eigenschapsbenadering dan kan worden geconstateerd dat ook hier een functie-ana-

lyse uitgangspunt vormt voor de keuze van de voorspellers. Binnen de eigenschapsbenadering zal men echter niet kunnen volstaan met het definiëren van effectief en ineffectief functiegedrag, maar zal tevens moeten worden aangegeven welke vaardigheden, capaciteiten en persoonlijkheidskenmerken bepalend zijn voor dit gedrag.

Deze stap van vereist functiegedrag naar psychologische eigenschappen is niet zonder problemen. Stereotype opvattingen over functies, impliciete persoonlijkheidstheorieën en onduidelijke definiëring van functie-inhoud en persoonskenmerken kunnen gemakkelijk leiden tot beoordelingsfouten (Greuter, 1991).

Functie-analyse methoden zijn er dan ook veelal op gericht het effect van dit soort fouten te verminderen. Een mogelijkheid is om een functie eerst op te splitsen in deeltaken, of nog specifiek in functie-elementen en vervolgens per deeltaak of functie-element te beoordelen welke eigenschappen daarvoor van belang zijn. Deze oordelen per deeltaak of element moeten daarna dan weer gecombineerd worden tot oordelen op het niveau van de totale functie. Als voorbeeld van methoden waarbij deze werkwijze wordt gevolgd noem ik de Position Analysis Questionnaire (PAQ), oorspronkelijk ontwikkeld door McCormick et al. (1972). Een andere mogelijkheid om de vertaalslag naar persoonseigenschappen te vergemakkelijken is door bij de beoordeling gebruik te maken van eigenschapsschalen waarbij concrete verichtingen of voorbeeldberoepen fungeren als schaalankers of referentiepunten. Hierbij kan bijvoorbeeld worden gedacht aan de Task Ability Scales van Fleishman c.s. (Fleishman, 1967; Fleishman en Quaintance, 1984).

Om vanuit de eigenschapsbenadering op een efficiënte manier tot goede voorspellingen te komen is het niet voldoende om te weten welke eigenschappen voor een bepaalde functie van belang zijn, maar zal ook bekend moeten zijn door welke tests of combinaties van tests deze eigenschappen worden gemeten.

Anders gezegd, er is informatie nodig over de constructvaliditeit van tests in relatie tot de te meten eigenschappen. Wat dit punt betreft kan worden geconstateerd dat in de jaren voorafgaand aan het artikel van Wernimont en Campbell nogal wat activiteiten waren ontwikkeld. De belangrijkste principes van het onderzoek naar de constructvaliditeit

van tests waren al in 1955 geformuleerd door Cronbach en Meehl. In de regelmatig verschijnende publicaties van Buros (vanaf 1938) werden tests beoordeeld en werd de relevante testinformatie samengevat. Ook het overzicht van Ghiselli vormde een bruikbare indeling. Tenslotte waren er daarnaast ook diverse, op factoranalytisch onderzoek gebaseerde indelingen van tests beschikbaar, waaronder de uitbreiding van het systeem van primaire factoren van Thurstone door French et al. (1963), de hiërarchische intelligentiemodellen van Vernon (1950, 1965) en, het meest spectaculair, Guilfords Structure of Intellect Model (Guilford, 1956, 1959, 1967). Dit laatste model fungeerde tot op zekere hoogte als een 'periodiek systeem' op grond waarvan zelfs nog niet als zodanig gemeten capaciteiten geprognosticeerd konden worden.

Waarom was de eigenschapsbenadering gegeven deze toch indrukwekkende systemen niet succesvoller wat betreft de predictieve validiteit van tests? Volgens Dunnette (1976) zaten de problemen vooral in de vertaalslag die moest worden gemaakt om van het ene systeem over te gaan naar het andere. De taxonomieën van functiekenmerken of functiegedrag en psychologische eigenschappen zouden met andere woorden onvoldoende op elkaar zijn afgestemd.



Het artikel van Wernimont en Campbell heeft nogal wat invloed gehad binnen de selectiepsychologie en het onderscheid 'signs versus samples' heeft inmiddels een eigen plaats gekregen in de literatuur op dit gebied (Dunnette, 1976; Schmitt en Ostroff, 1986; Robertson en Smith, 1989). Wat Nederlandse auteurs betreft noem ik nogmaals Roe die, o.a. in zijn hoofdstuk over personeelsselectie in het Nieuw Handboek Arbeids- en Organisationspsychologie (Roe, 1990) aangeeft dat de eigenschapsbenadering en de gedragsgeneralisatie-benadering fundamenteel van elkaar verschillen, in die zin dat er verschillende wetenschappelijke voorspellingsprincipes aan ten grondslag liggen. Voor een bespreking, en uitwerking van deze principes verwijs ik naar het betreffende hoofdstuk.

Wat ik hier wel wil laten zien en wat meteen ook kan dienen als indeling van mijn verdere verhaal is een ingekorte versie van een overzicht dat Roe geeft van predictoren ingedeeld naar voorspellingsprincipe, of zeg maar naar type benadering.

Tabel 1. *Overzicht van predictoren (naar Roe, 1990).*

Eigenschapsbenadering	Gedragsgeneralisatie-benadering
<i>Tests/Proeven</i>	
Capaciteitentests	Arbeidsproeven
Persoonlijkheidstests	Situatietests
<i>Interviews</i>	
Eigenschap-georiënteerd Interview	Criterium-georiënteerd (situatieel) interview
<i>Assessment Center</i>	

In de tabel worden drie typen predictoren onderscheiden, nl. tests/-proeven, interviews, en als laatste het assessment center. In het volledige overzicht worden daarnaast ook nog beoordelingen en biografische technieken genoemd als categorieën, maar die laat ik hier nu buiten beschouwing.

Zoals te zien zijn binnen de categorieën uitwerkingen naar beide voorspellingsprincipes mogelijk (op het assessment center kom ik nog apart terug).

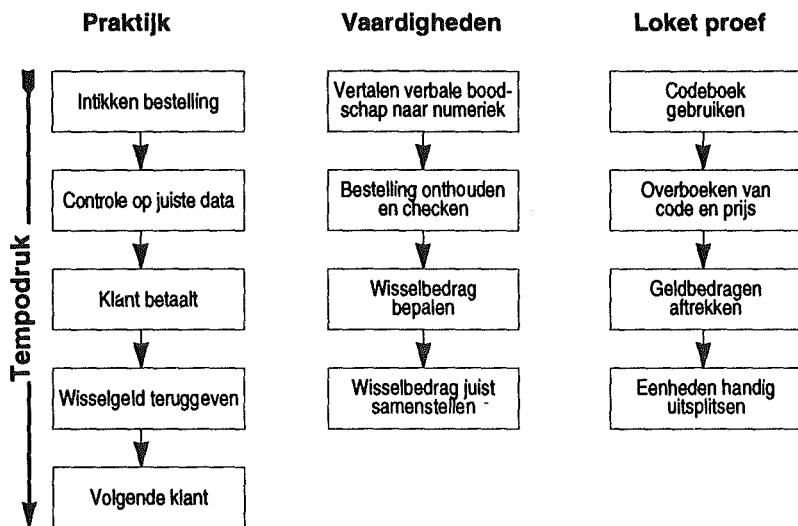
Arbeidsproeven

Kijken we naar de eerste groep dan vormen de arbeidsproeven, die je zou kunnen afzetten tegen capaciteitentests de meest duidelijke illustratie van de domein-steekproef-benadering. Ik ga er daarom relatief uitgebreid op in. Kort gezegd gaat het bij deze instrumenten om nabootsingen van (delen van) concrete functies. Aan kandidaten wordt gevraagd een aantal praktische vaardigheden te demonstreren die ook in de beoogde functie voorkomen. Dus aan een sollicitant voor de functie van administratief medewerker zou men kunnen vragen een standaardtekst in de computer in te voeren, of een daarin al aanwezige tekst te corrigeren. Aan een aspirant kok zou gevraagd kunnen worden m.b.v. een aantal ingrediënten een zo goed mogelijke maaltijd te bereiden of om een aantal recepten en menu's te beoordelen. En aan een hoogleraar in spé zou men kunnen opdragen een proefcollege te geven of om een onderzoeksopzet te bedenken of te becritiseren. Dat doet toch aan als een heel redelijke en natuurlijke vorm van selecteren. De kandidaat moet gewoon laten zien dat hij dat kan waar hij voor wil worden aangenomen. Over allerlei abstracte eigenschappen als verbale aanleg, analytisch vermogen, etc. hoeven we geen uitspraken te doen. Het is dan ook niet verwonderlijk dat de acceptatie van dit soort selectiemiddelen relatief hoog is. Daar komt nog bij dat zorgvuldig geconstrueerde arbeidsproeven bij de voorspelling van latere arbeidsprestaties over het geheel genomen zeer goede resultaten blijken op te leveren, met validiteits-coëfficiënten in dezelfde orde van grootte als die welke gevonden worden voor combinaties van capaciteitentests (Hunter en Hunter, 1984).

Ook uit eigen onderzoek komt dit naar voren. Ik denk bijvoorbeeld aan een onderzoek (al weer een aantal jaren geleden) i.v.m. de constructie en ijking van een Nederlandse testserie voor applicatieprogrammeurs (van der Flier, Keers en Drenth, 1976). Eén van de subtests, Diagrammen, had vrij sterk het karakter van een arbeidsproef. Per opgave werd een probleemstelling geformuleerd en werd een

stroomdiagram gegeven waarin d.m.v. cellen en pijlen de oplossing van het probleem stap voor stap werd uitgebeeld. In deze stroomdiagrammen moest steeds een aantal lege cellen worden ingevuld. Het onderzoek wees uit dat deze test van alle subtests de hoogste predictieve validiteit had m.b.t. cursusresultaten en praktijkcriteria en dat de validiteit ongeveer even hoog was als die van een optimale combinatie van een verbale, een numerieke en een ruimtelijk inzicht test.

Een ander voorbeeld betreft een onderzoek door NS-collega Schoonman (Schoonman en van der Flier, 1991) die een arbeidsproef voor lokettisten ontwikkelde of nauwkeuriger gezegd, een eerder door Kouwijzer (1971) ontwikkelde proef moderniseerde. Deze test, de loketproef, is een papieren simulatie van de verkoop van plaatsbewijzen door loketpersoneel. De opbouw van de proef is gebaseerd op een gedetailleerde analyse van de taak. Figuur 1. geeft naast elkaar een schematische representatie van de taakcyclus, onderverdeeld in elkaar opvolgende handelingen, de operationele vaardigheden die i.v.m. deze handelingen van belang zijn en de vertaling hiervan naar elementen van de arbeidsproef.



Figuur 1. De taakcyclus van de lokettist vanuit drie invalshoeken (Schoonman en Van der Flier, 1991).

De proef omvat de volgende deeltaken: De testkandidaat leest de opdracht van een reiziger. M.b.v. een codeboek wordt de bestelling vertaald naar een code waarmee vertrekstation, station van aankomst, soort plaatsbewijs en het al of niet in aanmerking komen voor korting wordt weergegeven. Daarna wordt de prijs opgezocht in een prijslijst en genoteerd. Het bedrag waarmee de reiziger betaald heeft staat erachter reeds vermeld. Als laatste moet het wisselbedrag en het wisselgeld worden genoteerd, waarbij de opdracht is het bedrag terug te geven in zo weinig mogelijk munt- en biljettenheden. Ook bij de scoring van de test is getracht een link te leggen met de praktijk. De relatieve ernst van verschillende typen fouten (afgeven van een verkeerd plaatsbewijs, teveel of te weinig wisselgeld teruggeven, onhandig omgaan met wisselgeld) komt tot uiting in het scoringsvoorschrift.



Gegevens over de predictieve validiteit van deze vernieuwde loketproef m.b.t. opleidings- en praktijkprestaties zijn nog niet bekend. Wel blijkt de proef die wordt afgenomen met een tijdslimiet van 12 minuten opmerkelijk hoog te correleren met de overall uitkomst van het psychologisch onderzoek dat ondermeer gebaseerd is op een testbatterij met 8 verschillende capaciteitentests.

Arbeidsproeven lijken wat betreft acceptatie en validiteit dus bijzonder aantrekkelijke selectiemiddelen te zijn. Toch kleven er ook een aantal nadelen aan. In de eerste plaats zijn ze relatief gevoelig voor verschillen in ervaring met de functie of met de betreffende taak.

Om even terug te keren naar het voorbeeld van de Diagrammentest, hoewel voor het maken van de opgaven geen programmeerkennis vereist is, kon wel worden geconstateerd dat voorafgaande ervaring in deze richting een positieve invloed heeft op de prestaties. Arbeidsproeven geven dus niet zozeer een beeld van wat kandidaten zullen kunnen leren als wel van wat ze al geleerd hebben. Dit betekent tevens dat bij de hoge validiteiten van arbeidsproeven m.b.t. opleidingsresultaten en praktijkcriteria de kanttekening moet worden geplaatst dat het verband met prestaties op langere termijn mogelijk minder sterk is dan men op grond van de relaties met korte termijn criteria zou verwachten. Het valideren van arbeidsproeven tegen arbeidsprestaties kort na indiensttreding kan dus gemakkelijk een te rooskleurig beeld opleveren.

Een tweede nadeel is dat arbeidsproeven gebonden zijn aan een specifieke functie. Dit impliceert dat voor verschillende functies of groepen functies ook verschillende arbeidsproeven vereist zijn. Zo wijkt bijvoorbeeld de kaartverkoop door conducteurs op de trein zodanig af van de kaartverkoop aan het loket dat een aparte kaartverkoopproef voor conducteurs ontwikkeld moest worden. En toen daarna in het kader van een adviesproject t.b.v. British Rail een 'ticket selling' test voor Engelse conducteurs ontwikkeld moest worden kon niet worden volstaan met een eenvoudige omzetting van de Nederlandse kaarverkoopproef maar moest weer een volledig nieuwe variant worden geconstrueerd, geënt op de Engelse situatie.

Consequentie van dit alles is dat de ontwikkelkosten van arbeidsproeven (die in het algemeen niet lager zullen zijn dan die van capaciteitentests, omdat wat betreft standaardisatie en betrouwbaarheid aan dezelfde eisen voldaan moet worden) veelal over minder grote aantallen afnames moeten worden afgeschreven. Daarnaast is ook het combineren van kandidaten voor verschillende functies bij groepsgebijze testafname minder goed mogelijk.

Een derde nadeel dat wat betreft oorzaken en consequenties overigens sterk samenhangt met het vorige is dat arbeidsproeven relatief snel

verouderen. Ik noemde dat al in verband met de noodzaak van modernisering van de loketproef, maar het geldt in feite ook voor het andere voorbeeld, de Diagrammentest die niet meer voldoende correspondeert met de wijze waarop applicatieprogrammeurs van nu hun werk doen. Dat wil niet zeggen dat de test absoluut niet meer valide is, maar wel dat de test als arbeidsproef minder goed voldoet.

Een vierde nadeel dat met name geldt voor complexe arbeidsproeven is dat bij sollicitanten die niet aan de eisen voldoen niet duidelijk is waar het nu precies aan ligt. De waarde van de arbeidsproef als diagnostisch instrument is m.a.w. beperkt. Natuurlijk kan het standpunt worden verdedigd dat dit bij selectie-onderzoek geen echt bezwaar is. Het gaat in deze visie bij selectie om het identificeren van geschikte sollicitanten en niet om het verklaren waarom bepaalde andere sollicitanten niet geschikt zijn. Anderzijds kan vanuit een ruimere visie op het selectie- en adviseringsproces, ook worden gepleit voor minder (uitsluitend) op de functie en meer (ook) op de mogelijkheden van de persoon gerichte feedback.

Op dit moment lijkt het zinvol kort stil te staan bij een al door Wernimont en Campbell genoemd voordeel van de gedragsgeneralisatiebenadering, namelijk dat deze benadering beter zou voldoen aan de eisen van non-discriminatie en 'equal opportunity'. Argumenten hiervoor waren dat de relevantie van de selectiemiddelen voor de inhoud van het werk gemakkelijker zou kunnen worden aangetoond dan bij de eigenschapsbenadering en dat ook het gevaar van een discriminerende testinhoud, die er toe zou kunnen leiden dat bepaalde groepen te lage testcores zouden behalen, minder groot zou zijn.

Inderdaad kan het bij arbeidsproeven als een voordeel worden beschouwd dat verschillen in prestaties tussen groepen en individuen met een verschillende culturele achtergrond niet worden geïnterpreteerd in termen van stabiele en voor een deel zelfs aangeboren eigenschappen. Anderzijds lijkt het oordeel over de grotere eerlijkheid van arbeidsproeven toch wat eenzijdig gebaseerd te zijn op wat binnen de Amerikaanse jurisprudentie op dit gebied aan argumenten wordt geaccepteerd.

De gevoeligheid van arbeidsproeven voor verschillen in ervaring met de taak of onderdelen daarvan maakt dat minderheidsgroepen die min-

der mogelijkheden hebben gehad relevante ervaring op te doen, relatief in het nadeel zijn. Om soortgelijke redenen werd i.v.m. de selectie voor voortgezet onderwijs in ontwikkelingslanden, in situaties waarin de kwaliteit van het voorafgaande onderwijs sterk varieerde gepleit voor gebruik van capaciteitentests als aanvulling op studietoetsen (Drenth, 1979). Ook voor capaciteitentests geldt overigens dat de relevantie voor een bepaalde functie heel goed aantoonbaar is indien men beschikt over een adequate functie-analyse-methode en er ook voldoende informatie beschikbaar is over de constructvaliditeit van de betreffende tests.

Tenslotte geldt dat er juist door het hanteren van het eigenschapsbegrip allerlei mogelijkheden ontstaan voor onderzoek naar (bronnen van) interne bias (zie b.v. Evers en van der Flier, 1992).

Persoonlijkheidstests

Ik wil het nu kort hebben over persoonlijkheidstests en met name over persoonlijkheidsvragenlijsten. In verband met dit type predictoren heeft in het verleden een in feite nog verdergaande discussie plaatsgevonden dan die van 'signs versus samples'. Deze ging niet zozeer over de vraag of we persoonlijkheidstrekken nodig hebben bij de voorspelling van toekomstige arbeidsprestaties als wel over de vraag of persoonlijkheidstrekken überhaupt bestaansrecht hebben, dus met andere woorden of gedrag niet zodanig door de situatie wordt bepaald dat er weinig reden is om cross-situationeel constante eigenschappen te veronderstellen. Tegenover trektheoretici als Eysenck (1960) en Cattell (1965) is met name door auteurs als Hettema (1967) en Mischel (1968) deze situationele bepaaldheid sterk benadrukt. Experimenten als die van Milgram (1974) waarin werd aangetoond dat heel normale proefpersonen mits in een daarop gerichte situatie gebracht tot moreel uiterst verwerpelijk gedrag in staat zijn en dat dergelijk gedrag nogal gemakkelijk aan de persoonlijkheid wordt toegeschreven leveren voor deze visie een aansprekend voorbeeld.

Feij (1985) komt, de bevindingen van diverse kanten op een rij zettend tot de conclusie dat de assumpties van de trek-theorieën weliswaar zijn gerelativeerd maar zeker niet verworpen. In zijn woorden staat 'Het oude en bijna triviale idee, dat vooral de interactie tussen type-

rende eigenschappen van mensen en situatievariabelen bepalend is voor individuele verschillen in gedrag weer centraal in de persoonlijkheidspsychologie'. Dit komt onder meer tot uiting in de belangstelling voor zogenaamde Situatie-Response (S-R)-vragenlijsten waarmee onderzocht kan worden in welke mate verschillen tussen mensen verklaard kunnen worden in termen van situaties, persoonlijkheidsfactoren en responsen en de (statistische) interacties daartussen (zie bv. Altink, Born en Algera, 1988).

De tegenstelling 'sign versus sample' die afkomstig lijkt te zijn van Goodenough (1949) werd i.v.m. persoonlijkheidsvragenlijsten ook al eerder naar voren gebracht door Cronbach (1960). Het ging daarbij om het probleem dat respondenten bij het invullen van persoonlijkheidsvragenlijsten in meerdere of mindere mate geneigd zullen zijn sociaal wenselijke antwoorden te geven. Zou de testconstructeur erop uit zijn om d.m.v. de vragenlijst een waarheidsgetrouw en representatief beeld te krijgen van gedrag en gevoelens van de respondenten, dan zou dit verschijnsel van sociale wenselijkheid de waarde van zijn instrument aanzienlijk beperken. De door Cronbach aangedragen oplossing voor dit probleem was dat het niet nodig is de antwoorden te interpreteren als betrouwbare zelfbeschrijvingen maar dat men ze eenvoudig kan zien als stukjes verbaal gedrag die op de een of andere manier indicatief zijn voor de te meten persoonlijkheidstrekk. Dus, om een door Cronbach gegeven voorbeeld aan te halen: wanneer neurotische respondenten op de vraag 'is uw gezondheid beter of slechter dan die van andere mensen van uw leeftijd' relatief vaak 'slechter' antwoorden dan kan dit item worden gebruikt om neuroticisme te meten, ook al is de gezondheid van neurotici gemiddeld niet slechter dan die van anderen. De empirische relatie met een extern criterium was dus bepalend voor het al of niet opnemen van een item in de vragenlijst of schaal.

Duidelijk zal zijn dat dit uitgangspunt voor testconstructie waarbij het steekproef- of inventarisatieprincipe is losgelaten aansluit bij de eerder besproken 'sign'-benadering, in die zin dat de betekenis die aan een instrument wordt toegekend minder stoelt op de concrete testinhoud dan op de relaties van de testcores met een extern criterium of met andere variabelen binnen een nomologisch netwerk.

Het was met name ook deze empirisch geöriënteerde 'sign'-benadering waar Wernimont en Campbell vraagtekens bij zetten.

Als voorspellers van latere arbeidsprestaties zijn persoonlijkheidsvragenlijsten aanzienlijk minder succesvol dan arbeidsproeven of capaciteitentests (Hunter en Hunter, 1984; Schmitt, Gooding, Noe en Kirsch, 1984). Toch worden zij bij selectie-onderzoek behoorlijk veel gebruikt. Een argument daarvoor kan zijn dat ze weinig overlap vertonen met andere onderdelen van het testprogramma waardoor ze toch een belangrijke toegevoegde waarde kunnen hebben. Een ander argument is dat ze geobjectiveerde informatie over persoonlijkheidskenmerken van kandidaten opleveren die als een check kan dienen op subjectieve indrukken o.g.v. het selectie-interview.

Over de te onderscheiden persoonlijkheidsdimensies lijkt de laatste jaren een vrij grote mate van overeenstemming te zijn ontstaan tussen verschillende onderzoekers op dit gebied. In de meeste gevallen wordt uitgegaan van vijf brede dimensies, bijgenaamd de 'Big Five' die kunnen worden aangeduid met de begrippen Extraversie, Emotionele stabiliteit, Vriendelijkheid, Zorgvuldigheid of Gewetensvolheid en Intellect of Openheid voor nieuwe ervaringen (Digman, 1990).

In een recent artikel geven Barrick en Mount (1991) een overzicht van de relaties van deze dimensies met verschillende typen prestatie-criteria voor verschillende beroepsgroepen. Volgens verwachting bleek de dimensie Zorgvuldigheid die ook kenmerken omvat als doelgericht, plichtsgetrouw en volhardend consistente relaties te hebben met alle prestatie-criteria voor alle beroepsgroepen. Voor de dimensie Emotionele stabiliteit werd dit, tegengesteld aan de verwachting, niet gevonden. Voor de overige dimensies bleken de relaties te variëren per beroepsgroep en per type criterium. Zo bleek Extraversie een valide predictor te zijn voor beroepen waarbij sociale interactie van belang is en de dimensie Intellect of Openheid voor nieuwe ervaringen voor succes binnen trainingsprogramma's.

Hoewel het hier gaat om betrekkelijk bescheiden correlaties blijkt deze indeling dus op zijn minst interessante aanknopingspunten op te leveren voor de selectiepraktijk. Recent onderzoek naar de structuur van Nederlandse eigenschapstermen waarin van een vergelijkbare in-

deling wordt uitgegaan (zie Hofstee en de Raad, 1991) onderstreept dit punt.

Interviews

Ook bij het interview zal ik niet te lang stilstaan. In het schema van Roe wordt onderscheid gemaakt tussen het eigenschap-georiënteerde interview en het criterium-georiënteerde interview. In het eerste geval gaat het erom, zoals daarnet al aangegeven, op basis van anamnestiche gegevens, observaties en gespreksindrukken te komen tot een beoordeling van de mate waarin de kandidaat bepaalde, voor de functie relevante persoonlijkheidseigenschappen bezit. In het tweede geval wordt getracht om op grond van in het verleden vertoond gedrag voorspellingen te doen over het te verwachten gedrag in de functie. Een variant op dit tweede type interview is het zgn. situationele interview (Latham, Saari, Pursell en Campion, 1980; Latham en Saari, 1984) waarbij aan kandidaten wordt gevraagd aan te geven hoe zij zich in bepaalde kritische situaties zouden gedragen en waarbij de antwoorden achteraf, zoveel mogelijk via vaste regels worden gescoord. Meta-analytische studies waarin de resultaten van grote aantallen onderzoeken worden gecombineerd hebben uitgewezen dat het criterium-gerichte interview over het geheel genomen een hogere validiteit heeft dan het eigenschaps-gerichte interview (Arvey en Campion, 1982). En ook voor het situationele interview geldt dat aanzienlijk hogere validiteiten zijn gevonden dan bij het meer gebruikelijke eigenschap-georiënteerde interview te verwachten zou zijn (Arvey, Miller, Gould en Burch, 1987).

Een van de redenen van deze hogere validiteiten kan zijn dat de interviewer bij de criterium-gerichte benadering meer gedwongen wordt zich af te vragen wat voor gedrag in de functie vereist is. Anderzijds zijn ook de sterkere structurering van het interview en de systematische scorings- of beoordelingsprocedure validiteitbevorderende factoren (Campion, Pursell en Brown, 1988; zie ook Drenth, 1988).

Jansen (1991) wijst er overigens terecht op dat het selectie-interview nog andere functies heeft dan het wegen van de kandidaat. Het fungeert als een persoonlijke ontmoeting, moet de interesse van de kandidaat in de functie versterken of in ieder geval niet onbedoeld verzwak-

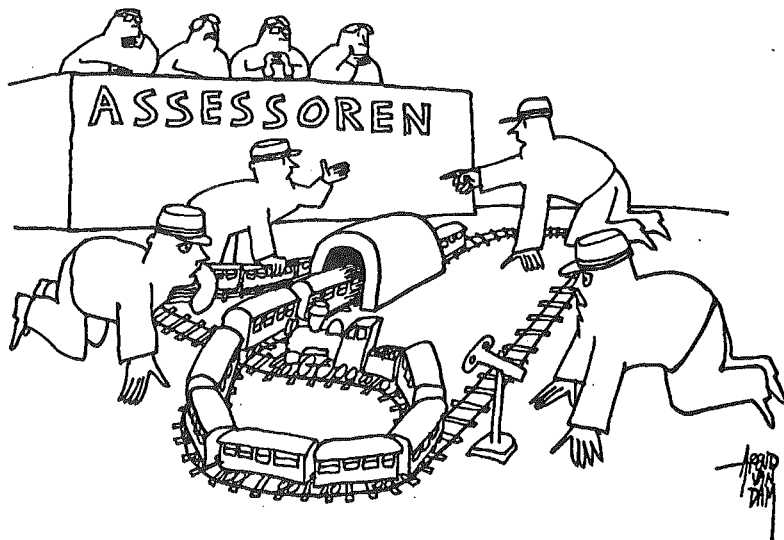
ken, het verschaft de kandidaat informatie over de toekomstige werkomgeving, etc. Dit impliceert dat de structurering van het interview niet onbeperkt kan worden opgevoerd en dat men zich in het gesprek niet al te strikt kan beperken tot voor de functie relevant gedrag. Anderzijds valt te verwachten dat een duidelijk aan functie-eisen gekoppelde gespreksinhoud de acceptatie van het interview als selectiemiddel in het algemeen zal bevorderen. Bovendien zal bij een duidelijk criterium-gericht interview bij kandidaten minder gauw het gevoel ontstaan dat met de vragen een inbreuk wordt gemaakt op de individuele privacy.

Assessment Centers

Ik wil het tenslotte hebben over de assessment center methode. Een assessment center kan worden omschreven als een procedure waarbij verschillende, elkaar aanvullende technieken voor de beoordeling van de geschiktheid van kandidaten op een geïntegreerde manier worden gehanteerd en waarbij verschillende beoordelaars of assessoren worden ingeschakeld. Wat de technieken betreft zijn vooral situatietests kenmerkend voor het assessment center. In deze tests worden belangrijke elementen van de betreffende functie nagebootst met de bedoeling gedrag op te roepen dat representatief is voor toekomstig functiegedrag. Situatietests komen qua uitgangspunt sterk overeen met arbeidsproeven, de scoring en interpretatie liggen echter minder vast (vgl. Roe, 1983). Voorbeelden van situatietests zijn de in-basket test of postbak (die ook wel bij de arbeidsproeven wordt gerekend - ik kom daar nog uitvoerig op terug), de groepsdiscussie, de analyse-/presentatie-oefening, de planningsoefening, het functioneringsgesprek (met een slecht functionerende medewerker) en het verkoopgesprek (uiteraard met een lastige klant). Naast situatietests kunnen nog allerlei andere technieken deel uitmaken van een assessment center, waaronder intelligentie- en persoonlijkheidstests en ook interviews.

In de meest zuivere vorm van het assessment center fungeren leidinggevenden uit de betreffende organisatie als assessoren. De assessoren dienen een goed beeld te hebben van de functie-eisen. Tevens hebben ze als het goed is een assessorentraining ondergaan, gericht op het observeren en beoordelen van gedrag in de diverse oefeningen. De asses-

soren zijn in principe volledig verantwoordelijk voor de uitkomsten van het assessment center. Het gemeenschappelijk eindoordeel per kandidaat komt tot stand o.g.v. een combinatie van de verschillende beoordelingen, veelal na onderlinge discussie. De rol van de psycholoog beperkt zich tot die van coördinator en procedurebewaker.



Wat betreft de predictieve waarde van het assessment center worden over het geheel genomen zeer positieve resultaten gemeld (Thornton en Byham, 1982; Hunter en Hunter, 1984; Schmitt, Gooding, Noe en Kirsch, 1984; Gaugler, Rosenthal, Thornton en Bentson, 1987). Hierbij worden echter wel een aantal vraagtekens geplaatst die vooral te maken hebben met het mogelijk optreden van meer of minder subtiële vormen van criteriumcontaminatie (in die zin dat het latere succes van kandidaten wordt beïnvloed door de uitkomsten van het assessment center). Ook wordt gewezen op het mogelijk effect van 'self-fulfilling prophecies' (Klimoski en Brickner, 1987).

In de indeling van predictoren naar de eigenschapsbenadering en de gedragsgeneralisatie-benadering wordt het assessment center gezien als een combinatie van de twee. De keuze van de oefeningen (simulaties van belangrijke taakelementen) sluit duidelijk aan bij de gedragsgeneralisatie-benadering. Het is ook duidelijk de bedoeling gedrag op

te roepen dat overeenkomt met gedrag dat in de functie van kandidaten gevraagd wordt.

Anderzijds vindt beoordeling en rapportage veelal plaats in termen van psychologische dimensies die het karakter hebben van eigenschappen, ook al zijn deze sterker gedefinieerd in gedragstermen en worden ze aangeduid met gedragsdimensies of gedragscriteria (vgl. Jansen, 1991). De structuur van een assessment center kan worden weergegeven in een schema zoals in Tabel 2 (waarin de combinaties worden aangeduid van oefeningen (praktijkopdrachten) en gedragsdimensies of -criteria waarop het in de oefeningen vertoonde gedrag zal worden beoordeeld.

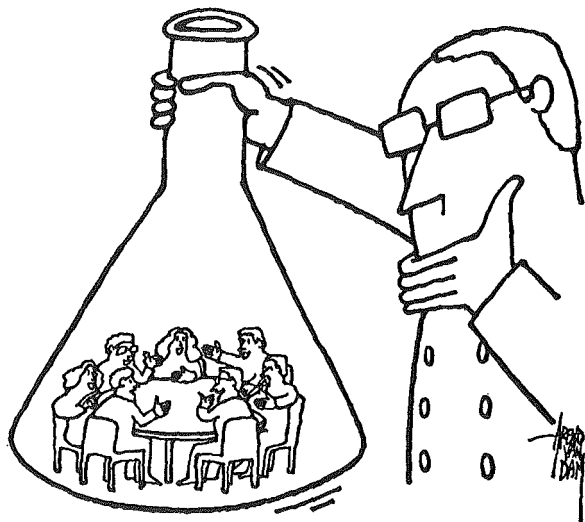
Tabel 2. *Matrix van management dimensies en praktijk opdrachten (Zaal, 1991).*

	IB	NO	GD	TW	PRES
Leiderschap	x		x	x	
Sensitiviteit	x		x	x	x
Planning & Organisatie	x				x
Delegatie & Controle	x			x	
Besluitvaardigheid	x		x	x	x
Inzicht	x	x	x	x	
Probleem Analyse	x	x	x	x	x
Mondelinge Communicatie			x	x	x
Schriftelijke Communicatie	x	x			

IB = In Basket, NO = Nota, GD = Groepsdiscussie, TW = Tweegesprek, PRES = Presentatie.

Het voorbeeld betreft een door Zaal (1991) beschreven assessment center gericht op managementfuncties bij de Rijksoverheid. Te zien is dat de verschillende opdrachten steeds informatie geven over maar een deel van de dimensies. Dus het schrijven van een nota zegt vooral wat over inzicht, probleemanalyse en schriftelijke communicatie, en een groepsdiscussie over leiderschap, sensitiviteit, besluitvaardigheid, inzicht, probleemanalyse en mondelinge communicatie. Bij de keuze van de oefeningen zal in het algemeen dan ook rekening gehouden worden met de dimensies die in de betreffende functie van belang zijn.

Bij een assessment center geldt veelal dat dimensies in meer dan één opdracht worden beoordeeld. Dat geeft de mogelijkheid om te onderzoeken of tussen beoordelingen op eenzelfde dimensie, verkregen via verschillende opdrachten, voldoende samenhang bestaat, in vergelijking met de samenhang tussen verschillende dimensies binnen eenzelfde opdracht en in vergelijking met de samenhang tussen verschillende dimensies gemeten via verschillende opdrachten. Dat kan op een systematische manier, via factoranalyse of m.b.v. een zgn. multi-methode multi-trek analyse (Campbell en Fiske, 1959). Ik ga daar nu niet op in. Van belang is dat zowel door Zaal als door diverse andere onderzoekers (Robertson, Gratton en Sharpley, 1987; Bycio, Alvares en Hahn, 1987) bij herhaling is gevonden dat verschillende dimensies binnen eenzelfde oefening sterk samenhangen, maar dat de samenhang tussen beoordelingen op eenzelfde dimensie, verkregen via verschillende oefeningen gering is. Er is dus maar weinig empirische steun voor het hanteren van algemene over situaties generaliseerbare gedragsdimensies. Dit resultaat is overigens niet toe te schrijven aan slordigheid van de beoordelaars, want die blijken het over het algemeen redelijk met elkaar eens te zijn.



SAMPLE VAN WERKGEDRAG: ASSESSMENT CENTRE METHODE...

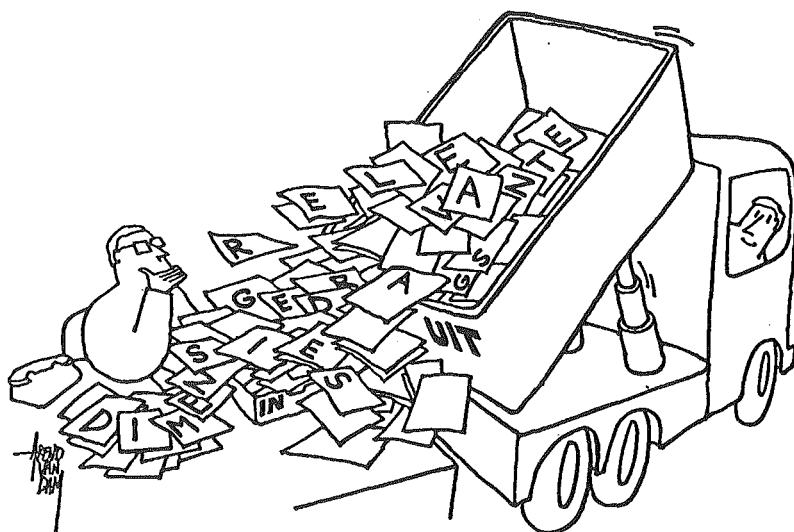
Zaal trekt hieruit de conclusie dat men voor de beschrijving van het managementprofiel van de kandidaten niet moet uitgaan van de managementdimensies als zodanig maar van de matrix van dimensies en opdrachten. De managementdimensies worden dus niet opgevat als algemene eigenschappen maar worden gezien als situatie-specifieke gedragsdimensies waarvan de betekenis gebonden is aan de opdrachten waarin zij worden waargenomen.

Dus we hebben het bij leiderschap over leiderschap binnen de context van de in-basket test of over leiderschap in de groepsdiscussie-situatie. Jansen (1991) komt in zijn boek over het beoordelen van managers o.g.v. dezelfde gegevens tot een wat andere conclusie. Hij stelt dat waar generalisering vanuit elementaire gedragsobservaties over situaties of oefeningen niet mogelijk blijkt, generalisering over dimensies binnen situaties geïndiceerd is. Concreet komt dit neer op het voorstel de dimensies maar helemaal te laten vallen en assessment center gedrag net als bij een proefwerk te beoordelen per situatie of oefening. Als middel voor het onderzoek van psychologische dimensies wordt het assessment center dus gewoon niet geschikt geacht.

Inderdaad is het moeten werken met situatie-specifieke gedragsdimensies of eigenschappen niet erg aantrekkelijk. Getracht kan echter worden, zoals Jansen trouwens ook aangeeft, de gedragsbasis van de eigenschappen te verbreden door er voor te zorgen dat er binnen een situatietest meer dan één meting van de eigenschappen plaatsvindt. Een goed voorbeeld van een situatietest waarbij dit mogelijk is en waarbij je zou kunnen spreken van een aantal parallel-oefeningen binnen één relatief brede context is de eerder genoemde postbak of in-basket test.

In-Basket

De in-basket wordt in de praktijk vaak als het belangrijkste onderdeel van het assessment center gezien en vormt in feite op zichzelf een mini-assessment center. Met de ontwikkeling van een in-basket voor managementfuncties hebben we de laatste jaren bij NS enige ervaring opgedaan. Ik wil daar kort wat over vertellen en probeer daarna een conclusie te formuleren.



IN - BASKET...

In een in-basket test wordt het postbakje met ingekomen post van een manager in een fictieve organisatie gesimuleerd. Het is de bedoeling dat de kandidaat de rol van de manager op zich neemt en binnen een bepaalde tijd schriftelijk reageert op de ingekomen post. Vaak wordt ook nog gevraagd de reactie te motiveren. De post, bestaande uit brieven, memo's, etc. betreft allerhande meer of minder urgente, maar meestal nogal lastige bedrijfsproblemen. Die problemen zijn natuurlijk niet maar willekeurig gekozen - ze komen als het goed is overeen met de problemen die ook werkelijk van belang zijn i.v.m. het al of niet slagen in de functie waarvoor de kandidaat wordt onderzocht.

Bij het ontwikkelen van een in-basket test voor managementfuncties bij NS werd begonnen met het opstellen van een lijst van relevante gedragsdimensies. Daarbij werd enerzijds geput uit de algemene managementliteratuur en anderzijds aangesloten bij de visie op management en leidinggeven die binnen het bedrijf was ontwikkeld (Cornwall, 1987). Vervolgens werd een 10-tal managers, afkomstig van verschillende bedrijfsonderdelen gevraagd de relevantie van de dimensies voor de eigen taak te beoordelen en bij de als meest relevant beoordeelde dimensies kritieke voorvallen te beschrijven of wel voorbeelden te geven van effectief of ineffectief gedrag. Deze kritieke

voorvallen vormden het basismateriaal voor de test. Ze werden vertaald naar problemen binnen een nieuwe fictieve organisatie en in de vorm gegoten van ingekomen post.

Na een vooronderzoek waarin de minder goed differentiërende items werden verwijderd en het aantal dimensies werd beperkt (Cornwall, 1988) bleef een test over met 13 items en 6 beoordelingsdimensies (zie Tabel 3).

Tabel 3. *Beoordelingsdimensies In Baskettest*

Analyseren van problemen/plannen

Sensitiviteit/zelfinzicht

Leidinggeven

Besluitvaardigheid

Visie

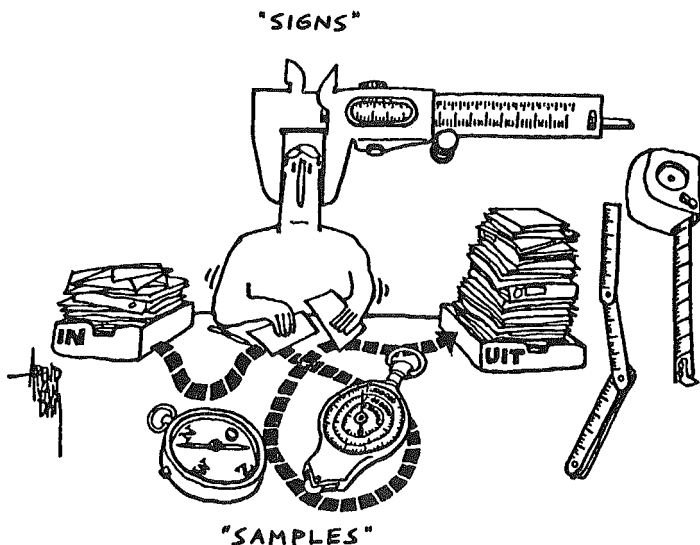
Management effectiviteit

Nu kun je de reacties in een test als deze op twee manieren beoordelen (zie voor een overzicht van de constructieproblematiek bij in-basket tests bijv. Born, Algera en Hoolwerf, 1988). Je kunt de antwoorden als geheel beoordelen op de verschillende dimensies, evt. met een paar voorbeelden van hoog en laag scorende reacties ernaast, of je kunt de reacties beoordelen aan de hand van een gedetailleerd beoordelingsvoorschrift waarbij eerst wordt nagegaan uit welke antwoordelementen de antwoorden zijn opgebouwd, waarna o.g.v. een van te voren ontwikkelde sleutel aan die elementen scores op de in aanmerking komende dimensies worden toegekend. Voordeel van de tweede manier is dat je veel objectiever kunt scoren, nadeel is dat je specifieke combinaties van antwoordelementen niet goed kunt honoreren. Beide manieren hebben we toegepast. Voor de details van het onderzoek verwijst ik naar een artikel van Bosch, van der Flier en Hoekstra dat is aangeboden aan Gedrag en Organisatie.

Interessant is wellicht nog te vermelden dat bij de ontwikkeling van het beoordelingsvoorschrift naast psychologen ook weer managers zijn ingeschakeld voor de beoordeling van antwoordelementen op algemene managementeffectiviteit. Wat leverde dit onderzoek op? Ik beperk me tot de voornaamste uitkomsten:

1. De interbeoordelaarbetrouwbaarheden blijken voor beide beoordelingswijzen hoog te zijn.
2. Bij de beoordeling volgens het beoordelingsvoorschrift is er aanzienlijk minder overlap tussen de dimensies dan bij beoordeling van antwoorden als geheel. In het laatste geval is er slechts sprake van één algemene factor.
3. De interne consistenties zijn in beide gevallen laag. Dit geldt zowel voor de specifieke dimensie-oordelen als voor de beoordeling op management effectiviteit.

Om afzonderlijke dimensies te kunnen onderscheiden is er dus op zijn minst een beoordelingsvoorschrift nodig. Wanneer beoordelaars de kans krijgen om hun eigen persoonlijke mening in te brengen blijft er maar één dimensie over die te interpreteren valt als het goed of slecht doen op de oefening. Daarnaast moet worden geconstateerd dat we bij de in-basket te maken hebben met een instrument dat qua homogeniteit dus zeg maar qua inter-item-betrouwbaarheid niet voldoet aan de gebruikelijke psychometrische eisen. Nu hebben we die homogeniteit in ons onderzoek aan een vrij rigoreuze test onderworpen: de antwoorden van respondenten werden uitgetypt en per item, steeds in een willekeurige volgorde aan de beoordelaars aangeboden. Ook in ander onderzoek zijn echter lage homogeniteiten gevonden voor in-basket tests (Brannick, Michaels en Baker, 1989; Shippmann, Prien en Katz, 1990). Wat dat betreft is het wellicht ook te vroeg voor een definitief oordeel over de mogelijkheden om met situatietests als de in-basket verschillende gedragsdimensies in beeld te krijgen en moet eerst nog verder worden onderzocht in hoeverre toepassing van de gangbare principes van test- en schaalconstructie tot betrouwbaarder maten kan leiden.

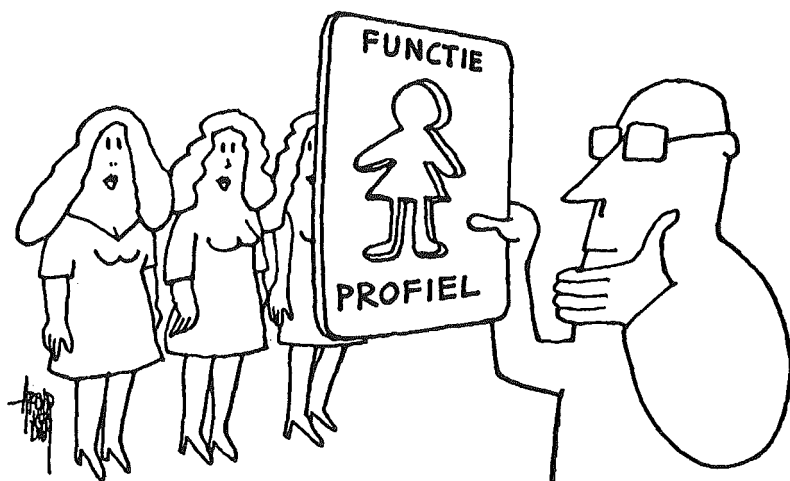


Conclusie

Dames en heren, het heeft mijns inziens nogal wat voordelen om in psychologisch selectie-onderzoek gebruik te kunnen maken van het eigenschapsbegrip. De uitspraak dat niets zo praktisch is als een goede theorie geldt zeker ook hier. Door uit te gaan van algemene psychologische eigenschappen kunnen op grond van een relatief klein aantal variabelen gedrag en prestaties in uiteenlopende functies worden voorspeld. Wijzigingen in functies behoeven niet direct te leiden tot de constructie van nieuwe predictoren en aan kandidaten die voor een bepaalde functie worden afgewezen kan, indien gewenst, informatie worden gegeven over hun verdere mogelijkheden. En ook wanneer zoals steeds vaker het geval is de advisering zich niet primair richt op een concrete startfunctie maar op een totale loopbaan biedt de eigenschapsbenadering duidelijke voordelen. Eigenschappen zijn dus nodig, en als het gaat om cognitieve capaciteiten ook bijzonder nuttig.

Vergelijkend onderzoek naar de utiliteit (het economisch nut) van verschillende selectiemiddelen (Hunter en Hunter, 1984) levert voor capaciteitentests de hoogste waarden op. Persoonlijkheidstests en het eigenschapsgerichte interview doen het wat validiteit en utiliteit betreft echter aanzienlijk minder goed zodat het weinig zin heeft om een algemene voorkeur uit te spreken voor de eigenschapsbenadering. Overigens blijken functie- of situatiespecifieke predictoren in bepaalde gevallen ook in vergelijking met capaciteitentests duidelijk hogere validiteiten op te leveren. Wat dat betreft is er tot op zekere hoogte sprake van een afruil tussen validiteit en generaliseerbaarheid. Een investering in functie-specifiek instrumentarium kan dus verantwoord zijn, bijvoorbeeld bij functies met een zeer groot afbreukrisico, bij functies die van strategisch belang worden geacht voor de organisatie of bij functies waarvoor zeer grote aantallen kandidaten moeten worden onderzocht.

In geval van combinaties van de twee benaderingen kan de keuze variëren per type instrument (b.v. capaciteitentests met een situationeel interview) of per fase in de selectieprocedure (b.v. voorselectie op grond van capaciteitentests en in de tweede fase een arbeidsproef, zie b.v. Jansen, 1986). Tenslotte zullen ook technologische ontwikkelingen, en met name de automatisering binnen de psychologische adviespraktijk de afweging beïnvloeden. Enerzijds ontstaan er nieuwe mogelijkheden voor het simuleren van taken (arbeidsproeven, situatie-tests). Anderzijds ontstaan meer mogelijkheden om via geautomatiseerde systemen m.b.v. in de computer opgeslagen expert-oordelen gedetailleerde functie-informatie om te zetten in profielen van functie-eisen waar eigenschapsprofielen van kandidaten tegen kunnen worden afgezet (capaciteitentests, persoonlijkheidsvragenlijsten). Daarmee wordt de vertaling van functiekenmerken naar psychologische eigenschappen, door Dunnette (1976) genoemd als zwakke schakel binnen de eigenschapsbenadering in ieder geval in technisch opzicht gemakkelijker uitvoerbaar.



Dames en heren, het onderscheid tussen 'signs' en samples' kan worden gezien als een relatieve kwestie. Tussen de interpretatie van gedrag in termen van algemene psychologische eigenschappen en de benadering waarbij trekken overbodig worden geacht bestaan tal van gradaties. Situatie-specifieke gedragsdimensies zijn daarvan een voorbeeld.

Assessment centers hebben als kenmerk dat ze assessoren richten op concreet gedrag als basis voor het doen van uitspraken en dat de eigenschappen waarover uitspraken worden gedaan zoveel mogelijk in gedragstermen worden gedefinieerd. Dit blijkt echter niet voldoende te zijn om deze gedragsdimensies ook daadwerkelijk van elkaar te kunnen onderscheiden.

Assessment centers zijn qua voorspellende waarde weliswaar succesvol, maar waar dit succes precies op gebaseerd is is onvoldoende duidelijk.

Onderzoek naar de condities waaronder de diagnostische waarde van assessment centers en situatietests als de in-basket kan worden vergroot is wetenschappelijk gezien interessant en van groot belang voor de selectie-praktijk.

Dankwoord

Dames en heren,

Aan het eind gekomen van deze oratie wil ik het Bestuur van de Vereniging voor Christelijk Wetenschappelijk Onderwijs, het College van Bestuur van de Vrije Universiteit en al degenen die aan mijn benoeming hebben bijgedragen danken voor het in mij gestelde vertrouwen. Ik beschouw het als een voorrecht verbonden te zijn aan deze bijzondere universiteit waar ik ook mijn wetenschappelijke opleiding heb mogen ontvangen.

Mijn dank gaat eveneens uit naar de Directie van de Nederlandse Spoorwegen en naar de leiding van de Service Eenheid Arbeidsomstandigheden die mij de gelegenheid hebben geboden deze part-time aanstelling aan de VU te aanvaarden.

Dames en heren medewerkers van de Service Eenheid Arbeidsomstandigheden,

Voor een arbeids- en organisatiepsycholoog bieden de Nederlandse Spoorwegen en de Streek- en Stadsvervoerbedrijven in Nederland een bijzonder interessant en uitdagend werkterrein. De combinatie van eisen wat betreft dienstverlening en veiligheid waaraan deze bedrijven moeten voldoen vraagt om gespecialiseerde ondersteuning op het gebied van de personeelsselectie en arbeidsomstandigheden. In deze behoefte proberen we te voorzien vanuit een multidisciplinair samengestelde Service Eenheid Arbeidsomstandigheden waarin de verschillende specialismen op ARBO-gebied bijeen zijn gebracht. Hierdoor kan over een breed terrein worden geadviseerd en kunnen problemen indien nodig vanuit verschillende invalshoeken worden geanalyseerd en aangepakt.

Binnen een dergelijke organisatie is een optimale afstemming tussen de verschillende disciplines voorwaarde voor succesvol functioneren

en vormt het bereiken van deze afstemming een belangrijke doelstelling die ook vanuit de ondersteunende research bevorderd dient te worden.

Ik heb de afgelopen jaren in verschillende functies met velen van u op een erg plezierige manier samengewerkt. Hopelijk zal dat in de toekomst ook mogelijk zijn. Werk genoeg in ieder geval, zeker voor het Kwaliteit Bewakingsbureau dat in samenwerking met de verschillende disciplines staat voor omvangrijke projecten als invoering van een werkplekgericht Periodiek Bedrijfsgeneeskundig Onderzoek, automatisering van het psychologisch testonderzoek, ontwikkeling van testprocedures t.b.v. allochtone sollicitanten en verdere ontwikkeling van instrumentarium voor de meting van taakkenmerken en werkbeleving. De NS-psychologen die al direct na mijn komst bij NS in 1983 voorspelden dat ik wel hoogleraar zou worden gun ik deze casus. Hun overige voorspellingen zal ik uiteraard kritisch blijven volgen.

Dames en heren hoogleraren en medewerkers van de Faculteit der Psychologie en Pedagogische Wetenschappen,

Ik stel het bijzonder op prijs in uw midden te mogen werken. Door mijn beperkte aanwezigheid zal een vrij groot deel van mijn universitaire activiteiten zich afspelen binnen de vakgroep en in CAOP-verband. Daarnaast ben ik echter graag bereid mij waar dat zinvol is, in te zetten voor samenwerking met andere vakgroepen. Voor enkele gezamenlijke projecten met de vakgroep Sociale Psychologie zijn inmiddels voorstellen ingediend.

Dames en heren medewerkers van de vakgroep Arbeids- en Organisatie-psychologie,

Onze banden dateren van lang geleden. Ik heb mij vanwege de open en collegiale sfeer en de belangstelling voor elkaars werk altijd bijzonder thuis gevoeld binnen de vakgroep. Ook na mijn vertrek naar NS in 1983 zijn de contacten blijven bestaan en hebben we op diverse manieren samengewerkt. Het afgelopen jaar zijn oude banden aangehaald en is rond nieuwe onderzoeksthema's samenwerking ontstaan. Binnen wat ik voor het gemak maar even de wetenschappelijke competitie noem, heeft de vakgroep zich in het verleden altijd gemanifesteerd als een bijzonder creatief en goed ingespeeld elftal. Naar mijn idee is dat

nog steeds zo. Dat met de AIO's het aantal spitsen relatief sterk is toegenomen maakt de wedstrijd volgens mij alleen maar interessanter.

Hooggeleerde Drenth, beste Pieter,

Een voordeel van dat ingespeelde elftal is dat de spelers elkaar niet de hele tijd in het oog hoeven te houden om goed samen te werken. Dat geldt denk ik ook voor ons. Van jou heb ik het vak geleerd, een vak dat dankzij jouw dubbele leerstoel voor mij tal van raakvlakken kreeg met cross-culturele psychologie en testontwikkeling t.b.v. onderwijs in ontwikkelingslanden. Ik beschouw het als een eer jou voor een deel te mogen vervangen, of in termen van het elftal als invaller voor jou te mogen fungeren. Ik verheug me bijzonder op de verder te ontwikkelen gezamenlijke activiteiten.

Dames en heren studenten in de Arbeids- en Organisatie-psychologie,
Voor de kwaliteit van het Wetenschappelijk Onderwijs is binnen deze Universiteit nog onlangs en met nadruk de aandacht gevraagd. U hebt, ook gezien de eisen die aan u worden gesteld, recht op goed onderwijs, met enthousiasme gegeven en professioneel georganiseerd. Ik wil aan dit onderwijs graag mijn bijdrage leveren en hoop op een positieve respons van uw kant.

Dames en heren,

Ik dank u voor uw aandacht.

Noot

De auteur dankt Marise Born en Wouter Schoonman voor hun kritisch commentaar op een eerdere versie van dit manuscript.

Literatuur

- Altink, W.M.M., Born, M.Ph. en Algera, J.A. (1988). De S-R-vragenlijstmethode bij organisatievraagstellingen. *Gedrag en Organisatie*, 1, 5, 38-53.
- Arvey, R.D. en Campion, J.E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology*, 35, 281-322.
- Arvey, R.D., Miller, H.E., Gould, R. en Burch, P. (1987). Interview validity for selecting sales clerks. *Personnel Psychology*, 40, 1-12.
- Barrick, M.R. en Mount, M.K. (1991). The Big Five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*, 44, 1-26.
- Born, M.Ph., Algera, J.A. en Hoolwerf, G. (1988). Dilemma's bij het meten van managementkwaliteiten. *Gedrag en Organisatie*, 1, 1, 30-46.
- Bosch, J., van der Flier, H. en Hoekstra, H.A. (1992). Twee beoordelingswijzen voor een In-basket test in een bedrijfssetting. Aangeboden aan: *Gedrag en Organisatie*.
- Brannick, M.T., Michaels, Ch.E. en Baker, D.P. (1989). Construct validity of in-basket scores. *Journal of Applied Psychology*, 74, 957-963.
- Bycio, P., Alvares, K.M. en Hahn, J. (1987). Situational specificity in assessment center ratings: a confirmatory factor analysis. *Journal of Applied Psychology*, 72, 463-474.
- Campbell, D.T. en Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Campion, M.A., Pursell, E.D. en Brown, B.K. (1988). Structured interviewing: Raising the Psychometric Properties of the Employment Interview. *Personnel Psychology*, 41, 25-42.
- Cattell, R.B. (1965). *The scientific analysis of personality*. Chicago: Aldine Publishing Company.
- Cornwall, N. (1987). In Basket: *Toepassing van Assessment Center-Methoden bij de selectie voor leidinggevende functies binnen de Nederlandse Spoorwegen*. Utrecht, N.V. Nederlandse Spoorwegen.
- Cornwall, N. (1988). *Ontwikkeling van een in-basket test ter bepaling van management-potentieel voor de Nederlandse Spoorwegen*. Doctoraal-werkstuk, Universiteit van Amsterdam. Utrecht: N.V. Nederlandse Spoorwegen.
- Cronbach, L.J. en Meehl, P.E. (1955). Construct validity in psychological

- tests. *Psychological Bulletin*, 52, 281-302.
- Cronbach, L.J. (1960). *Essentials of psychological testing*. New York: Harper en Row.
- Digman, J.M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417-440.
- Drenth, P.J.D. (1979). Prediction of school performance in developing nations: schoolgrades or psychological tests? *Journal of Cross-Cultural Psychology*, 8, 49-70.
- Drenth, P.J.D. (1988). De waarde van het selectie-interview. *Gedrag en Organisatie*, 1, 2, 18-26.
- Dunnette, M.D. (1976). Aptitudes, abilities and skills. In: M.D. Dunnette (Ed.). *Handbook of Industrial and Organizational Psychology*. Chicago: Rand McNally.
- Evers, A. en van der Flier, H. (1992). Alloctonen op de arbeidsmarkt. In: P.J.D.Drenth, Hk. Thierry en Ch.J. de Wolff (Red.). *Nieuw Handboek Arbeids- en Organisationspsychologie*. Deventer: van Loghum Slaterus.
- Eysenck, H.J. (1960). *The structure of human personality*. London: Routledge and Kegan Paul.
- Feij, J.A. (1985). Persoonlijkheidspsychologie. In: J.F. Orlebeke, P.J.D. Drenth, R.H.C. Janssen en C. Sanders (Red.). *Compendium van de psychologie*. Muiderberg: Coutinho.
- Fleishman, E.A. (1967). Development of a behavior taxonomy for describing human tasks: a correlational-experimental approach. *Journal of Applied Psychology*, 51.
- Fleishman, E.A. en Quaintance, M.K. (1984). Taxonomies of Human Performance. *The description of Human Tasks*. New York: Academic Press.
- van der Flier, H., Keers, W. en Drenth, P.J.D. (1976). *Applicatie programmeur test; constructie en ijking van een Nederlandse testserie voor applicatie programmeurs*. Lisse: Swets en Zeitlinger.
- French, J.W., Ekstrom, R.B. en Price, L.A. (1963). *Kit of reference tests for cognitive factors*. Princeton, N.J.: Educational Testing Service.
- Gaugler, B.B., Rosenthal, D.B., Thornton, G.C. en Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72, 493-511.
- Ghiselli, E.E. (1966). *The validity of occupational aptitude tests*. New York: Wiley.

- Goodenough, F. (1949). *Mental testing: Its history, principles and applications*. New York: Holt, Rinehart en Winston.
- Greuter, M.A.M. (1991). Arbeidsanalyse ten behoeve van personeelsselectie. In: J.A. Algera (red.). *Analyse van arbeid vanuit verschillende perspectieven*. Amsterdam/Lisse: Swets en Zeitlinger.
- Guilford, J.P. (1956). The structure of intellect. *Psychological Bulletin*, 53, 267-293.
- Guilford, J.P. (1959). Three faces of intellect. *American Psychologist*, 14, 469-479.
- Guilford, J.P. (1967). *The nature of human intelligence*. New York: Mc Graw-Hill.
- Hettema, P.J. (1967). Trekken, processen en persoonlijkheidstests. *Nederlands Tijdschrift voor de Psychologie*, 22, 618-641.
- Hofstee, W.K.B. en de Raad, B. (1991). Persoonlijkheidsstructuur: de AB5C-taxonomie van Nederlandse eigenschapstermen. *Nederlands Tijdschrift voor de Psychologie*, 46, 262-274.
- Hunter, J.E. en Hunter, R.F. (1984). Validity and utility of some alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Jansen, P.G.W. (1986). De twee fasen van psychologische selectie. In: J.C.J. Bonarius, W.Th.A.M. Everaerd en M.N. Verbaten (Red.). *Psychologie in Nederland - 1985. Deel II: Praktijgericht onderzoek*. Lisse: Swets en Zeitlinger.
- Jansen, P.G.W. (1991). Het beoordelen van managers: *Effectiviteit van assessment center methoden bij selectie en ontwikkeling van managers*. Baarn: Nelissen.
- Klimoski, R. en Brickner, M. (1987). Why do assessment centres work? The puzzle of assessment centre validity. *Personnel Psychology*, 40, 243-260.
- Kouwijzer, G.F. (1971) *Exploratief onderzoek naar taak en de selectie van lokettistes bij de N.V. Nederlandse Spoorwegen*. Utrecht: Subfaculteit Psychologie RUU/Nederlandse Spoorwegen.
- Latham, G.P., Saari, M., Pursell, E.D. en Champion, M.A. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422-427.
- Latham, G.P. en Saari, M. (1984). Do people do what they say? Further studies on the situational interview. *Journal of Applied Psychology*, 69, 569-573.

- McCormick, E.J., Jeanneret, R.R. en Mecham, R.L. (1972). A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire. *Journal of Applied Psychology Monograph*, 56, 347-368.
- Milgram, S. (1974). *Obedience to Authority*. New York: Harper en Row.
- Mischel, W. (1968). *Personality and assessment*. New York: Wiley.
- Robertson, I.T., Gratton, L. en Sharpley, D. (1987). The psychometric properties and design of managerial assessment centres: Dimensions into exercises won't go. *Journal of Applied Psychology*, 60, 187-195.
- Robertson, I.T. en Smith, M. (1989). Personnel Selection Methods. In: M. Smith en I.T. Robertson (Eds.). *Advances in Selection and Assessment*. New York: Wiley.
- Roe, R.A. (1983). *Grondslagen der personeelsselectie*. Assen: van Gorcum.
- Roe, R.A. (1990). Personeelsselectie: modellen en instrumenten. In: P.J.D. Drenth, Hk. Thierry en Ch. J. de Wolff (Red.). *Nieuw Handboek Arbeids- en Organisationspsychologie*. Deventer: van Loghum Slaterus.
- Schmitt, N., Gooding, R.Z., Noe, R.A. en Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37, 407-422.
- Schmitt, N. en Ostroff, C. (1986). Operationalizing the behavioral consistency approach: selection test development based on a content-oriented strategy. *Personnel Psychology*, 39, 91-108.
- Schoonman, W. en van der Flier, H. (1991). Arbeidsproeven als selectiemiddel: Drie voorbeelden. In: H. van der Flier, P.G.W. Jansen en J.N. Zaal (Red.). *Selectieresearch in de praktijk*. Amsterdam/Lisse: Swets en Zeitlinger.
- Shippmann, J.S., Prien, E.P. en Katz, J.A. (1990). Reliability and validity of in-basket performance measures. *Personnel Psychology*, 43, 837-859.
- Thornton, G.C. en Byham, W.C. (1982). *Assessment centres and managerial performance*. New York: Academic Press.
- Vernon, P.E. (1950). *The structure of human abilities*. London: Methuen.
- Vernon, P.E. (1965). Ability factors and environmental inferences. *American Psychologist*, 20, 723-733.
- Wernimont, P.F. en Campbell, J.P. (1968). Signs, samples and criteria. *Journal of Applied Psychology*, 52, 372-376.
- Zaal, J.N. (1991). Toepassingen van Assessment Centers. In: H. van der Flier, P.G.W. Jansen en J.N. Zaal (Red.). *Selectieresearch in de praktijk*. Amsterdam/Lisse: Swets en Zeitlinger.

ISBN 90 - 5383 - 018 - 5



VU Boekhandel / Uitgeverij
Hoofdgebouw Vrije Universiteit
De Boelelaan 1105, 1081 HV Amsterdam
Postbus 7161, 1007 MC Amsterdam
Telefoon 020 - 444 355, Telex 18191 vuboe nl